ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION SYSTEM AND

METHOD UTILIZING WHITENING FILTER

FOR UPDATING OF PREDICTOR COEFFICIENTS

5   ## CROSS REFERENCE TO RELATED APPLICATIONS

[001]   The present application claims the benefit of priority from U.S. Provisional

Patent Application No. 60/183,280, entitled "Adaptive Differential Pulse Code

Modulation System and Method Utilizing Whitening Filter For Updating Of

Predictor Coefficients" filed on February 17, 2000, which is incorporated by

10   reference herein.

BACKGROUND

Field of Invention

[002]   The present invention relates generally to encoding and decoding of

15   digital audio signals, and more particularly to predictor adaptation in adaptive

differential pulse code modulation (ADPCM) systems.

Description of the Prior Art

[003]   FIG. 1 may be referenced in conjunction with the following discussion.

ADPCM is a well-known technique for encoding speech and other audio signals

20   for subsequent transmission over a network. A standard implementation of such

a system is described in the International Telecommunication Union (ITU-T)

Recommendation G.722, 7 kHz Audio-Coding Within 64 kBit/s, which is

incorporated by reference herein.

[004]   As described in U. S. Pat. No. 4,317,208, issued February 23, 1982 to

Araseki et al. and incorporated by reference herein, a differential pulse code

5   modulation system is a band compression system in which a prediction of each

signal sample at a present time period is based on signal samples at past time

periods. Such a process is particularly effective with voice and similar band

signals due to their high degree of correlation between successive signal samples.

A predicted signal S_j at a time j is typically derived at a transmitter 102 by the

10   general equation:

$$S_j = A_1 S_{j-1} + A_2 S_{j-2} + \ldots A_n S_{j-n} ;$$

[005]   where A_1, A_2, ... A_n are termed the prediction coefficients. The prediction

coefficients are selected to minimize the difference between an input signal Y_j

and the predicted signal S_j, thus minimizing a prediction error E_j which is in turn

15   quantized and transmitted to a receiver 104, thereby requiring significantly less

transmission bandwidth than would the input signal. The receiver 104 works in

a manner generally the reverse of the transmitter 102, thereby reconstructing the

input signal.

[006]   The characteristics of a voice or related audio signal vary with time,

20   consequently the optimum coefficient values also vary. One method of

attempting to efficiently derive prediction coefficients is to adapt them with the

goal of minimizing the prediction error E_j while such error is being observed,

which could generally describe an ADPCM system. A common type of predictor employed in these systems is a pole-based predictor, such as predictors 110 and 126, which utilizes a feedback loop to minimize the energy in the prediction error signal E_j, which is sometimes referred to as the difference or residual signal.

5   [007]   Due to the reality of frequent transmission errors between the transmitter 102 and the receiver 104, the prediction errors $\hat{E}_j$ (which have been inverse quantized) produced at the receiver 104, and thus the reconstructed input signal $\acute{S}_j$ depending thereon, has a tendency to diverge from the real input signal Y_j received at the transmitter 102. To gradually eliminate the adverse effect of the

10   transmission errors, the prediction coefficients are typically derived by the general equation:

$$A_i^{j+1} = A_i^j (1-\delta) + g \cdot F_1(S'_{j-i}) \cdot F_2(\hat{E}_j);$$

[008]   where j=1 to n, δ is a positive value much smaller than 1, g is a proper positive constant, $\acute{S}_{j-i}$ is a reconstructed input signal delayed i samples, and F_1

15   and F_2 are non-decreasing functions. The receiver 104 prediction coefficient values are tracked, or gradually caused to converge to those of the transmitter 102, by operation of the term (1-δ). The detrimental effect of transmission errors is thus partially overcome.

[009]   Instability or oscillation of the receiver may still occur in pole-based

20   predictor systems due to the feedback loop to the predictor, which uses the prediction error signal $\hat{E}_j$ and the preceding reconstructed input signal $\acute{S}_{j-i}$ to derive the prediction coefficients as described above. Stability checking is often

3

used to ensure that the prediction coefficients remain in desired ranges, but at the expense of increased complexity as the number of poles, i.e., coefficients, increases.

[0010] In U. S. Pat. No. 4,317,208, Araseki et al. describe a system that also employs zero-based predictors, such as predictors 120 and 128, which do not utilize a feedback loop but which are known to provide less predictor gain than pole-based predictors and consequently inhibit or slow down the adaptation process. They propose that such a combination of pole-based and zero-based predictors may overcome the instability issues described above, and gain the advantages of each type of predictor.

[0011] In U. S. Pat. No. 4,593,398, issued June 3, 1986 to Millar and incorporated by reference herein, it is suggested that a pole-based predictor, even coupled with a zero-based predictor, is still vulnerable to mistracking if the input signal contains two tones of equal amplitude but different frequencies. Millar notes that certain input signals may cause the pole-based predictor adaptation driven by the feedback loop to have multiple stable states, thus the receiver 104 may stabilize with its prediction coefficients at values different than the transmitter 102. This in turn is likely to cause a distorted frequency response at the receiver 104 and its associated audio output device.

[0012] The Millar patent proposes to mitigate the problems associated with lower predictor gain in zero-based predictors and mistracking in pole-based predictors. The system described by Millar and depicted in FIG. 1 is such that the predictor

4

means in the transmitter 102 and the receiver 104 derive the prediction

coefficients based on an algorithm including a non-linear function having no

arguments comprising the value of the reconstructed input signal, such as signals

$Ś_j$ and $Ś_{j-i}$. This coefficient adaptation is depicted by arrows 119 and 127. This is

5   in contrast to the Araseki system wherein the prediction coefficients are partially

derived from a reconstructed input signal such as signal $Ś_{j-i}$, which is dependent

upon the predicted signal S_j, which is dependent upon all of the immediate past

coefficient values.

[0013] It is postulated that the Millar system and method may be

10   computationally expensive to implement. Therefore what is needed is a system

and methods in which the convergence to the optimal prediction coefficients,

and thus to the predicted signal S_j, occurs more rapidly and efficiently than in

prior art systems.

SUMMARY

[0014] An improved adaptive differential pulse code modulation (ADPCM) system and method comprises an encoder and a decoder linked together by a network connection and configured for processing digital audio signals. More

5 particularly, the technique described is related to adaptation of predictor coefficients in an ADPCM environment. The components of the system may be implemented in software form as instructions executable by a processor or in hardware form as digital circuitry. Furthermore, devices implementing the system and method described are preferably configured to include both an

10 encoder and a decoder for bi-directional communication with a similarly situated remote device, or may be configured with solely the encoder or decoder.

[0015] At the encoder, a digitized input signal is applied to a subtractor, which derives a difference signal by subtracting from the input signal a predicted signal generated by a pole-based predictor. After quantizing, transmitting to a decoder,

15 and inverse quantizing, the difference signal is added to the predicted signal by an adder to provide a reconstructed input signal, which is fed back to the predictor and to the subtractor. The encoder is additionally provided with a whitening filter for receiving the reconstructed input signal and applying thereto a filtering algorithm to generate a filtered reconstructed signal. The filtered

20 reconstructed signal is utilized to update, or adapt, the prediction coefficients of the pole-based predictor, thus providing more rapid and computationally efficient convergence to optimal prediction coefficients.

6

[0016] The decoder operates in an inverse manner to the encoder, receiving the quantized difference signal from an encoder and processing it to reconstruct the input signal for delivery to sound reproducing means. It is noted that devices employing the ADPCM techniques described herein are interoperable with devices employing prior art techniques, for example, those described in ITU-T G.722. It is further noted that the techniques described herein may be adapted for various implementations, one example being the employment of a plurality of encoders and/or decoders for frequency sub-band processing.

[0017] Other embodiments of the invention comprise additional predictors at the encoder and the decoder, operating to maximize the signal-to-noise ratio for certain input signals. The additional predictors are preferably zero-based predictors, the output therefrom being summed with the pole-based predictor output to produce the predicted signal.

BRIEF DESCRIPTION OF THE FIGURES

[0018] In the accompanying drawings:

[0019] FIG. 1 depicts a prior art ADPCM system;

[0020] FIG. 2 depicts an ADPCM system, in accordance with a first embodiment

5   of the invention; and

[0021] FIG. 3 depicts another ADPCM system, in accordance with a second

embodiment of the invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

[0022] FIG. 2 depicts a first embodiment of an ADPCM system 200 in accordance with the invention. ADPCM system 200 comprises an encoder 202 and decoder 204 linked in communication by a network connection 206, such as an ISDN line,

5 fractional T1 line, digital satellite link, wireless modems, or like digital transmission service. At encoder 202, a digitized input signal, typically representative of speech, is applied to a conventional subtractor 208. The input signal is represented as Y_j, signifying a value at sample period j. Subtractor 208 derives a difference signal E_j by subtracting from input signal Y_j a predicted

10 signal S_j generated by a pole-based predictor 210. The difference signal E_j is quantized by a conventional quantizer 212 to obtain a quantized numerical representation N_j for transmission to decoder 204 over the network connection 206. Quantizer 112 is preferably of the adaptive type, but a quantizer utilizing fixed step sizes may also be used.

15 [0023] Numerical representation N_j is also applied to a conventional inverse quantizer 214, which derives a regenerated difference signal D_j. A conventional adder 216 adds regenerated difference signal D_j to a predicted signal S_j (output by the pole-based predictor 210) to provide a reconstructed input signal X_j. The reconstructed input signal X_j is in turn applied to the pole-based predictor 210,

20 which calculates the predicted signal S_j in accordance with the following equation:

$$S_j = a_1^j S_{j-1} + a_2^j S_{j-2} + \ldots + a_n^j S_{j-n}$$

where S_{j-1} is a stored value of the predicted signal at sample period j-1, S_{j-2} is a stored value of the predicted signal at sample period j-2, and so on, and a_1^j to a_n^j are the predictor coefficients at sample period j, where n corresponds to the total number of poles (i.e., coefficients) of pole-based predictor 210. In one implementation of ADPCM system 200, the pole-based predictor 210 is limited to two poles, yielding the relation:

$$S_j = a_1^j S_{j-1} + a_2^j S_{j-2}.$$

[0024] The predicted signal S_j generated by predictor 210 is then applied to adder 216, completing the feedback loop.

[0025] Predictor coefficients a_1^j and a_2^j are updated in accordance with the generalized equations:

$$a_1^{j+1} = a_1^j (1 - \delta_1) + g_1 \cdot F_1(X_j^f, X_{j-1}^f, X_{j-2}^f)$$

$$a_2^{j+1} = a_2^j (1 - \delta_2) + g_2 \cdot F_2(X_j^f, X_{j-1}^f, X_{j-2}^f, X_{j-3}^f, a_1^j)$$

where X_j^f is a filtered version of reconstructed input signal X_j at sample period j; δ_1, δ_2, g_1 and g_2 are proper positive constants, and F_1 and F_2 are nonlinear functions which may consist of correlations, sign-correlations, or other relationships. Calculation of the filtered reconstructed signal X_j^f is discussed below.

[0026] In general, whitening filters modify the spectrum of signals to provide a flatter signal spectrum, so that there is less variation of energy as a function of

10

frequency. It is noted that a perfect white noise signal has equal energy at every frequency. Stochastic gradient adaptive filters generally converge more rapidly with white signals than with non-white signals. Therefore, the use of a whitening filter in the present system and method is preferred at least for its

5   effect on convergence of the adaptive pole-based predictors 210 and 226.

[0027] Referring back to FIG. 2, a whitening filter 218 receives the reconstructed input signal X_j and applies thereto a filtering algorithm to generate a filtered reconstructed signal X^f_j. To ensure stable operation of whitening filter 218, the filter coefficients $a_2^{f^{j+1}}$ and $a_1^{f^{j+1}}$ undergo the clamping set forth below at every

10   other time step (i.e., for odd values of j):

[0028] $a_2^{f^{j+1}}$ is clamped to a maximum of 12288 and a minimum of –12288; and

$$a_1^{f^{j+1}} \text{ is clamped in magnitude to } 15360 - a_2^{f^{j+1}}.$$

Implementation of this clamping routine is exemplified as:

$$\text{temp} = 15360 - a_2^{f^{j+1}};$$

15    $$\text{if } a_1^{f^{j+1}} > \text{temp, then } a_1^{f^{j+1}} \text{ is set to equal temp;}$$

$$\text{if } a_1^{f^{j+1}} < -\text{temp, then } a_1^{f^{j+1}} \text{ is set to equal } -\text{temp.}$$

[0029] The filtered reconstructed signal X^f_j output by whitening filter 218 is utilized to update the predictor coefficients $a_1{}^{j+1}$ and $a_2{}^{j+1}$, as described above and indicated on FIG. 2 by arrow 220.

20   [0030] According to a preferred implementation, whitening filter 218 has two zeroes, yielding the relation:

11

$$X_j^f = X_j - a_1^f X_{j-1} - a_2^f X_{j-2}$$

where a_1^f and a_2^f are the first and second order filter coefficients. The filter coefficients a_1^f and a_2^f are updated at each time step j in accordance with the following equations:

5
$$a_2^{f\,j+1} = a_2^{f\,j}\left(1-\left(\frac{256}{32768}\right)\right)-\left(\frac{1}{32}\right)a_1^{f\,j}\,\text{sgn}\left[X_j^f\right]\text{sgn}\left[X_{j-1}^f\right]+128*\text{sgn}\left[X_j^f\right]\text{sgn}\left[X_{j-2}^f\right];\ \text{and}$$

$$a_1^{f\,j+1} = a_1^{f\,j}\left(1-\left(\frac{128}{32768}\right)\right)+192*\text{sgn}\left[X_j^f\right]\text{sgn}\left[X_{j-1}^f\right];$$

where sgn [ ] is the sign function that returns a value of 1 for a nonnegative argument and a value of -1 for a negative argument.

[0031] In accordance with a computationally economical implementation of

10   ADPCM system 200, the values of the predictor coefficients may be frozen at every other sample interval j. It should be noted that by recalculating predictor coefficients for pole-based predictor 210 only at every other interval, computational resources are conserved. This implementation is described by the following equations:

15   for even j:

$$a_2^{j+1} = a_2^j;\ \text{and}$$

$$a_1^{j+1} = a_1^j;$$

else for odd j:

12

$$a_2^{j+1} = a_2^{j-1}\left(1 - \left(\frac{510}{32768}\right)\right) - \left(\frac{1016}{32768}\right)\lim\left[a_1^{j-1}\right]\text{sgn}\left[X_{j-1}^f\right]\text{sgn}\left[X_{j-2}^f\right] + 127*\text{sgn}\left[X_{j-1}^f\right]\text{sgn}\left[X_{j-3}^f\right]$$

$$-\left(\frac{1}{32}\right)\lim\left[a_1^{j-1}\right]\text{sgn}\left[X_j^f\right]\text{sgn}\left[X_{j-1}^f\right] + 128*\text{sgn}\left[X_j^f\right]\text{sgn}\left[X_{j-2}^f\right]; \text{ and}$$

$$a_1^{j+1} = a_1^{j-1}\left(1 - \left(\frac{127.5}{32768}\right)\right) + 191.25*\text{sgn}\left[X_{j-1}^f\right]\text{sgn}\left[X_{j-2}^f\right] + 192*\text{sgn}\left[X_j^f\right]\text{sgn}\left[X_{j-1}^f\right];$$

where sgn [ ] is the sign function that returns a value of 1 for a nonnegative

argument and a value of -1 for a negative argument, and

$$\lim\left[a_i^{j-1}\right] = a_i^{j-1} \text{ for } -8192 \le a_i^{j-1} \le 8191;$$

$$\lim\left[a_i^{j-1}\right] = -8192 \text{ for } a_i^{j-1} < -8192; \text{ and}$$

$$\lim\left[a_i^{j-1}\right] = 8191 \text{ for } a_i^{j-1} > 8191.$$

[0032] To ensure stability, a_2^{j+1} and a_1^{j+1} are clamped similarly to $a_i^{f\,j+1}$ and $a_i^{f\,j+1}$ as

described above. That is:

a_2^{j+1} is clamped to a maximum of 12288 and a minimum of -12288; and

a_1^{j+1} is clamped in magnitude to $15360 - a_2^{j+1}$.

Implementation of this clamping routine is exemplified as:

temp = 15360 - a_2^{j+1};

if a_1^{j+1} > temp, then a_1^{j+1} is set to equal temp;

if a_1^{j+1} < -temp, then a_1^{j+1} is set to equal -temp.

[0033] Decoder 204 operates in an inverse manner to encoder 202. Inverse

quantizer 222 receives the numerical representation N_j over network connection

206 and derives the regenerated difference signal D_j. Adder 224 sums the

regenerated difference signal D_j with the predicted signal S_j generated by pole-based predictor 226 to produce the reconstructed input signal X_j. The reconstructed input signal X_j is then delivered to sound-reproducing means (which will typically include a D/A converter and loudspeaker) for reproduction

5   of the speech represented by the input signal Y_j.

[0034] At the decoder 204, the reconstructed input signal X_j is additionally applied to whitening filter 230 and pole-based predictor 226. Pole-based predictor 226 operates in a substantially identical manner to pole-based predictor 210 of encoder 202 and generates as output predicted signal S_j, which is applied

10   to adder 224 to complete the feedback loop. Whitening filter 230, which operates in a substantially identical manner to whitening filter 218 of encoder 202, provides as output a filtered reconstructed signal X^f_j for use by pole-based predictor 226 in updating the predictor coefficients, as discussed above and indicated on FIG. 2 by arrow 228.

15   [0035] Those skilled in the art will recognize that the various components of encoder 202 and decoder 204 will typically be implemented in software form as program instructions executable by a general purpose processor. Alternatively, one or more components of encoder 202 and/or decoder 204 may be implemented in hardware form as digital circuitry.

20   [0036] Additionally, those skilled in the art will recognize that, although the pole-based predictors 210 and 226 are described above in terms of a two-pole

implementation, the invention is not limited thereto and may be implemented in connection with pole-based predictors having any number of poles.

[0037] It is additionally noted that the ADPCM technique embodied in the invention may be adapted in various well-known ways in order to improve the speed and performance of the encoding and decoding processes. For example, a transmitting entity may break the input signal into a plurality of frequency-limited sub-bands, wherein each sub-band is applied to a separate encoder operating in a substantially identical manner to encoder 202. The sub-banded encoded signals are then multiplexed for transmission to a receiving entity over the network connection. The receiving entity then demultiplexes the received signal into a plurality of sub-banded signals and directs each sub-banded signal to a separate decoder operating in a manner substantially identical to decoder 204. The sub-banded reconstructed signals are thereafter combined and conveyed to sound-reproducing means.

[0038] In other embodiments of the invention, additional predictors may be combined with the pole-based predictors to maximize the signal-to-noise ratio for certain input signals. Referring now to the FIG. 3 embodiment of an ADPCM system 300, encoder 302 differs from encoder 202 of the FIG. 2 embodiment by the addition of a conventional zero-based predictor 306. Zero-based predictor 306 receives the regenerated difference signal D_j and produces a zero-based partial predicted signal S_{jz}, which is added to the partial pole-based predicted signal S_{jp} (equal to S_j in the FIG. 2 embodiment) by adder 308 to provide

predicted signal S_j. Predicted signal S_j is in turn applied to the feedback loop of pole-based predictor 210 and to subtractor 208. It is noted that zero-based predictor 306 does not have a feedback loop, and its predictor coefficients are conventionally updated with dependence on regenerated difference signal D_j.

5   [0039] Similarly, decoder 304 differs from decoder 204 of the FIG. 2 embodiment by the inclusion of zero-based predictor 310. The regenerated difference signal D_j is applied to zero-based predictor 310, which generates as output a zero-based partial predicted signal S_{jz}. Adder 312 combines the zero-based partial predicted signal S_{jz} with pole-based partial predicted signal S_{jp} provided by pole-based

10   predictor 226 to produce the predicted signal S_j.

[0040] Another embodiment of the invention utilizes at least one look-up table in determining the proper coefficients for the predictors, i.e., pole-based predictors 210 and 226 of FIGs. 1 and 2, and/or zero-based predictors 306 and 310 of FIG. 3. For example, the first pole-based predictor coefficient is a function of three

15   quantities: its former value, the sign of the current value of the sum of the quantized prediction error plus the all-zero predictor, and the sign of the past value of the sum of the quantized prediction error plus the all-zero predictor. In this embodiment, no arithmetic is necessary in determining a prediction coefficient value, however, identical input-output characteristics of the predictors

20   are preserved.

[0041] It should be appreciated that devices utilizing the above-described ADPCM techniques, such as audioconferencing or videoconferencing endpoints,

will typically be equipped for bi-directional communications over the network connection, and so will be provided with both an encoder (such as encoder 202 or 302) for encoding local audio for transmission to a remote endpoint as well as a decoder (such as decoder 204 or 304) for decoding audio signals received from

5   the remote endpoint.

[0042] It is further noted that devices employing the above-described ADPCM techniques of the invention are advantageously interoperable with devices employing some prior art ADPCM techniques, such as those described in the aforementioned Millar reference and the ITU-T G.722 reference.

10   [0043] Finally, it is generally noted that while the invention has been particularly shown and described with reference to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention.

15